GENOMIC SEQUENCING OF HUMAN CHROMOSOME 19 AND COMPARATIVE ANALYSIS OF HUMAN AND RODENT DNA REPAIR GENE REGIONS.

Jane Lamerdin, Mishelle Montgomery, Stephanie Stilwagen, Melissa Ramirez, Aaron Adamson, Subha Basu, Ami Kyle, Paula McCready, Jeff Gingrich, Anne Olsen, Larry Thompson, Emilio Garcia, and Anthony Carrano. Human Genome Center, Biology and Biotechnology Research Program, Lawrence Livermore National Laboratory, Livermore, CA, 94550

Approximately 83% of human chromosome 19 is spanned by cosmids for which an EcoR1 map has been derived. We are scaling our sequencing facility to take advantage of these ordered clones to provide high-throughput, high-accuracy sequence for all of chromosome 19. In addition, we are utilizing our rapid clone selection/mapping capabilities to sequence other genomic regions associated with DNA repair and disease susceptibility. We have completed ~ 650 kbp of genomic sequence to high accuracy and have another ~500 kbp in various stages of completion (from assembly through annotation). Our primary effort has been targeted to cosmids containing the human DNA repair genes *HHR23A*, *XRCC1* and *ERCC2* on chromosome 19, *ERCC4* on chromosome 16, *XRCC3* on chromosome 14, and *XRCC2* on chromosome 7, as well as selected rodent homologs. We are also working on chromosome 19 regions associated with olfactory receptors and a congenital nephrotic disease.

We have sequenced 76 kbp containing the human and mouse *XRCC1* genes, which span 26 kbp in the mouse and 31.9 kbp in the human. In addition to the coding regions, 9 conserved elements were identified with sequence identities ranging from 65% to 78%. We have completed 54 kbp of human sequence encompassing the *ERCC2* gene as well as 54 kbp spanning the syntenic regions in the mouse and hamster. A defect in *ERCC2* leads to the cancer-prone human disorder xeroderma pigmentosum (XP-D). The human *ERCC2* gene is comprised of 23 exons and is 98% identical to the rodent homologs at the protein level. We identified two genes flanking *ERCC2*,. One may be a new member of the kinesin light chain gene family; the other has no known function. All three genes, and their orientation are conserved in the three mammals.

Like *ERCC2*, the *ERCC4* gene product is involved in the nucleotide excision repair pathway, which recognizes and removes DNA damage. The genomic region indicates that the full-length gene spans ~29 kbp and is >50% AT-rich. The *ERCC4* gene product exhibits significant homology to the *S. cerevisiae* rad1 and *S. pombe* rad16 genes, which encode single strand endonucleases.

We have sequenced a cosmid and its associated cDNA for the recently cloned human *XRCC3* gene, which appears to play a crucial role in chromosomal stability. The predicted protein shares residue identity with the GTP binding domain of the *S. cerevisiae* rad51 and rad57 proteins involved in recombinational repair. Sequence analysis of several candidate cDNAs for the XRCC2 gene also show similarity to the same domain in these proteins. Sequence analysis of the XRCC3-containing cosmid identified a second kinesin light chain gene physically linked to a DNA repair gene.

(This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under contract no. W-7405-ENG-48.)